



CHICAGO JOURNALS

---

The Quality of Writing Tasks and Students' Use of Academic Language in Spanish

Author(s): Amy C. Crosson, Lindsay Clare Matsumura, Richard Correnti, and Anna Arlotta-Guerrero

Source: *The Elementary School Journal*, Vol. 112, No. 3, Academic Language (March 2012), pp. 469-496

Published by: [The University of Chicago Press](#)

Stable URL: <http://www.jstor.org/stable/10.1086/663298>

Accessed: 05/10/2015 13:39

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



*The University of Chicago Press* is collaborating with JSTOR to digitize, preserve and extend access to *The Elementary School Journal*.

<http://www.jstor.org>

---

# THE QUALITY OF WRITING TASKS AND STUDENTS' USE OF ACADEMIC LANGUAGE IN SPANISH

---

## ABSTRACT

This study investigates the quality of the writing tasks assigned to native Spanish speakers in bilingual (Spanish-English) contexts, and the relationship between task quality and students' use of an academic register in their native language. Fifty-six language arts tasks were collected from 26 grade 4 and 5 teachers, and four student writing samples were collected in response to each task ( $N = 224$ ). Multilevel modeling revealed that variation in students' use of key features of academic language in their writing was associated with the cognitive demand of writing tasks. Findings suggest that students' opportunities to respond to challenging tasks when writing in their native language are rare and that the rigor of writing tasks may relate to students' production and development of academic language.

Amy C. Crosson  
Lindsay Clare Matsumura  
Richard Correnti  
Anna Arlotta-Guerrero  
UNIVERSITY OF  
PITTSBURGH

---

**T**HE ability to write well is a foundational skill for school success and beyond. However, research indicates that many schools are not providing quality instruction in writing, resulting in what is widely considered a “writing proficiency crisis” (Graham & Perin, 2007; National Commission on Writing, 2003). While recent results from the National Assessment of Educational Progress (NAEP) indicate that students' writing has improved on average, certain subgroups of students continue to lag far behind national norms (Salahu-Din, Persky, & Miller, 2008). Notably, only 17% of grade 8 students who are eligible for free or reduced-price lunch and 5% of English language learning (ELL) students scored at or above the proficient level in writing, compared to nearly one-third of grade 8 stu-

dents on the whole (Salahu-Din et al., 2008). Approximately three-quarters of ELL students are Latino (Goldenberg, 2008); thus, the majority of ELLs who struggle with academic writing are native speakers of Spanish.

It is increasingly recognized that many ELL students have difficulty with English-language writing tasks in part because of a lack of familiarity with the lexical, grammatical, and discursive features that are associated with *academic language* (Scarcella, 2002, 2003; Schleppegrell, 1996, 2004; Schleppegrell, Achugar, & Oteiza, 2004; Snow & Uccelli, 2009). Also referred to as *academic English* (Bailey, 2007; Scarcella, 2003) and *advanced literacy* (Christie, 2002), academic language relates to the forms and functions of language necessary for participation in contexts of schooling. The purpose of this study is to examine the opportunities to develop academic writing skills that are provided to native Spanish speakers in bilingual (Spanish-English) contexts. Specifically, we investigate the relationship between students' use of academic language when writing in Spanish and the rigor of Spanish-language writing tasks assigned to them.

### Academic Language in Spanish: A Potential Resource for Bilingual Students

While difficulty with academic language has been documented among ELL students writing in English, less clear is whether minority language students have developed an academic voice in their native language (L1). This question is of interest above and beyond the social and economic benefits of bilingualism and biliteracy, as there is evidence that many L1 literacy skills are advantageous to literacy development in English (Dressler & Kamil, 2006; Genesee, Lindholm-Leary, Saunders, & Christian, 2005). The underlying premise of Cummins's (1979) seminal and oft-cited developmental interdependence hypothesis, broadly interpreted, holds that L1 competencies will transfer to support the same types of cognitively demanding tasks in the target language once a threshold of target language proficiency has been achieved. In fact, scholars of academic language (Colombi & Schleppegrell, 2002; Scarcella, 2002; Schleppegrell, 2004) and leading practitioners (Teachers of English to Speakers of Other Languages [TESOL], 2009) have asserted that a command of academic language in students' L1 contributes to development of academic language in English.<sup>1</sup> And, research contrasting the lexical and grammatical features of academic language in English and Spanish has revealed that the linguistic resources are highly similar in the two languages (Gibbons, 1999; Gibbons & Lascar, 1998), further suggesting that, for those students who have developed proficiency with academic Spanish, there is strong potential for transfer of the features of academic language in Spanish to support academic writing in English.

To date, however, we lack empirical evidence of whether and how native Spanish speakers draw on their L1 resources to support academic writing in English. Thus while it is possible that native Spanish-speaking children who struggle with academic writing tasks in English have developed academic language in Spanish but are unable to transfer those skills into English, it is also possible that these children have not had sufficient opportunity to develop academic language in Spanish (Colombi, 2002; Colombi & Roca, 2003; Gibbons & Lascar, 1998). In this article, we investigate to what degree native Spanish speakers develop academic language in Spanish and the opportunities that expand or constrain use of academic language in their L1.

## Academic Language and Its Associated Linguistic Features

This study draws on the theoretical framework of Systemic Functional Linguistics (SFL; Halliday, 1994; Schleppegrell, 2004) as well as the developmental perspective on academic language offered by Snow and Uccelli (2009) to identify a set of lexical and grammatical features associated with academic language. These features are studied as indicators of an academic voice.<sup>3</sup>

Scholars in the SFL tradition situate all analysis of language within its purpose and context. As such, academic language is a “register” that both shapes and is shaped by the social contexts of its use (Schleppegrell, 2004). First, in contrast to informal registers used in face-to-face exchanges about everyday topics, academic language is typically used to communicate about abstract knowledge. Second, distinct from informal registers that are oral, interactive, and spontaneous, the academic register tends to be written, nonparticipatory, and logically organized (Schleppegrell, 2001). Third, this register usually assumes unequal relations between interactants, projecting an authoritative voice and low affective involvement (Christie, 2002; Eggins, 2004; Schleppegrell, 2001, 2004).<sup>3</sup>

Numerous SFL analyses have revealed how the academic register is realized through a co-occurring set of lexical and grammatical features (Colombi & Schleppegrell, 2002; Halliday, 1994; Schleppegrell, 2004; see also Eggins, 2004, for analytic method). For example, academic writing is characterized by a “prestige” or “specialist lexis” (Schleppegrell, 2004) that enables precision of meaning, often necessary for clear communication with a noninteractive audience and frequently essential when writing about abstract topics outside of everyday experience (Schleppegrell, 2004). Beck and McKeown (2001) and others (Bailey, Huang, Shin, Farnsworth, & Butler, 2007; Beck, McKeown, & Kucan, 2002; Nation, 2001) draw a useful distinction between specific, sophisticated vocabulary words that are generally useful across a range of academic contexts (e.g., *enormous* and *accustomed*) and those that are specialized and domain specific, representing technical meanings in particular academic contexts (e.g., *pollination* and *metamorphosis*). We call these *general academic vocabulary* and *specialized academic vocabulary*, respectively. Use of general and specialized academic vocabulary contributes to an academic voice by signaling an authoritative stance and, in some cases, association with an academic community, as is illustrated, for example, by the potentially important differences in the meanings of *large* versus *enormous* or *change* versus *metamorphosis*.

Because condensation of information is characteristic of academic language, grammatical structures that enable a writer to succinctly convey information and ideas are frequently employed in academic writing (Achugar, 2003; Colombi, 2002; Schleppegrell, 2004). SFL analyses demonstrate how the grammatical resource of embedding enables writers to condense information and present ideas and information efficiently (Schleppegrell, 2004). For example, embedded clauses enable writers to add information to noun groups (Fang, Schleppegrell, & Cox, 2006), as in *No sabía nada de sus hermanos que estaban en la guerra* (He didn’t know what happened to his brothers that were in the war). Use of embedding not only enables the writer to present information concisely, but it also gives the writer options for how to structure the text to connect and/or foreground ideas.

Our article is also informed by the theoretical framework for understanding academic language offered by Snow and Uccelli (2009)—itself influenced by SFL—that

attends specifically to the developmental challenges of learning to write in an academic register. Situating the challenges of learning to use an academic register within the development of communicative competencies beginning in early childhood, Snow and Uccelli identified some lexical and grammatical features of an academic register that are particularly relevant to studying young adolescent learners.

First, Snow and Uccelli (2009) noted that an important resource for clarifying relationships between text ideas and for promoting the logical organization of text is the use of a variety of connectives (i.e., conjunctions, adverbial phrases, and discourse markers that link ideas between clauses and sentences).<sup>4</sup> In particular, writers move toward a more academic register when they signal how information is organized through use of temporal connectives indicating logical relations or chronological relations with *primero*, *segundo*, *finalmente* (first, second, finally) (Halliday, 1994). Academic writing is also characterized by subordination of ideas through causal connectives such as *porque* (because).<sup>5</sup> Adversative connectives such as *aunque* (although) and *sin embargo* (however) play an important role in the rhetorical strategy of refuting counterarguments in persuasive writing (Crowhurst, 1990). In sum, connectives are a resource for developing academic writing, because instead of relying on the contextualized cues of face-to-face interaction, they make explicit the relationships between text ideas. Finally, Snow and Uccelli noted that high lexical diversity is characteristic of academic language; thus as word choice becomes more varied, it moves closer to approximating an academic register.

Of course, the value of any of these features of academic language lies not in their presence at a surface level but instead in the underlying discourse practices that the academic register helps to instantiate—for example, the realization of the appropriate stance on the part of the writer. Thus in the present study we analyzed students' use of lexical and grammatical features of academic language as indicators of an academic voice.

## Writing Tasks and Opportunities to Use Academic Language

While there have been important advances toward identifying the lexical and grammatical features of academic language, there is a pressing need to understand how instruction supports students' meaningful use of an academic register, whether in Spanish or English (Snow & Uccelli, 2009). Although carefully orchestrated classroom discussions certainly contribute to students' opportunity to learn to use academic language (e.g., Reznitskaya et al., 2001), many structures and rhetorical features common in academic writing are rare in oral language exchanges. Indeed, learning this register in any language requires opportunities to engage in writing in response to academic texts.

Converging evidence demonstrates that the quality of writing tasks is associated with the quality of students' writing (Clare & Aschbacher, 2001; Matsumura, Patthey-Chavez, Valdés, & Garnier, 2002; Monte-Sano, 2008) as well as gains on standardized tests of achievement (Carbonaro & Gamoran, 2002; Matsumura, Garnier, Pascal, & Valdés, 2002; Matsumura, Garnier, Slater, & Boston, 2008; Newmann, Bryk, & Nagaoka, 2001). Newmann and colleagues (2001), for example, found that students in grades 3, 6, and 8 who had been exposed to more rigorous writing tasks requiring interpretation, analysis, and evaluation were more likely to show greater gains on the Iowa Test of Basic Skills than those who had been exposed to poor-

quality writing tasks, controlling for students' prior achievement and demographics. Notably, students with both high and low prior achievement appeared to benefit from exposure to high-quality tasks. Matsumura, Garnier, et al. (2002) found that students in grades 7 and 10 whose teachers provided them with more rigorous tasks showed greater gains on the reading subscales of the SAT-9 than their peers in classrooms who were assigned lower-quality tasks.

While these studies provide some converging evidence of a critical link between the quality of writing tasks and student achievement, definitions of *achievement* and *writing quality* have been operationalized along broad or holistic dimensions. How tasks support (or fail to support) students' facility with specific features of writing—that is, features associated with an academic register—has remained unexamined, in spite of evidence that the writing of many students (especially ELL and low-income students) is less likely to reflect lexical, grammatical, and discursive features of academic language (Scarcella, 2003; Schleppegrell, 1996; Snow & Uccelli, 2009).

## Research Questions

This study investigated the use of academic language in the writing of fourth- and fifth-grade native-Spanish-speaking students in their L1 and the quality of the writing tasks they are assigned. Multilevel modeling techniques were used to examine the amount of variation in students' use of academic language at three levels: teachers, tasks, and students. The primary purpose of these analyses was to understand how variation in students' use of academic language was distributed across these levels, in particular between tasks within teachers since it is seldom that we have the opportunity to investigate a single molar unit of instruction as contributing to variance in student outcomes. If variation existed largely between students, we might infer that use of academic language in writing was an individual-level phenomenon, with students of higher ability likely to use a more academic voice. However, if variation were greater at the task or teacher level, this would suggest that there were indeed instructional effects in the data, where some tasks (and perhaps some teachers) were more likely to inspire use of an academic register. We further investigated instructional effects by examining whether specific characteristics of writing tasks predicted students' use of academic language.

The research questions we addressed were as follows: (1) To what degree do students use salient features of academic language when writing in their native language (Spanish)? (2) What is the quality of writing tasks assigned to students such that they are challenged to engage with rigorous texts, interpret texts, and use evidence to support assertions? (3) Does students' use of academic language vary as a function of individual teachers, tasks, and students? If so, does the quality of tasks predict students' use of salient features of academic language?

## Method

### Participants

The study was located in a medium-size urban district in the Southwest. The elementary schools ( $N = 12$ ) in this study served primarily Latino students from low-income families. On average, 76.1% of the population was classified as Hispanic,

and 93.8% were eligible for free and/or reduced-price lunch. Twenty-six fourth- and fifth-grade teachers who provided language arts instruction in Spanish participated in the study. Teachers' level of experience ranged from 1 year (six teachers) to 24 years (one teacher). Nearly half ( $n = 12$ ) had been teaching for less than 3 years. Eight teachers held emergency or provisional certification, and five held master's degrees. One teacher was National Board certified.

The students in this study were native speakers of Spanish enrolled in bilingual education classes in which classroom activities were carried out in both Spanish and English. In participating classrooms, 90% of the students were administered the Texas Assessment of Knowledge and Skills (TAKS) in Spanish. Students are only administered this exam in Spanish if they have more fully developed academic skills in Spanish than in English. Moreover, 95% of students in participating classrooms were designated as limited English proficient (LEP), demonstrating that their language competencies were stronger in Spanish than English.

### Procedure

The assignments and student work were collected over a 2-year period<sup>6</sup> as part of a study investigating relationships between literacy coaching, instructional quality, and student learning. Teachers were asked to submit four writing tasks that required students to respond to a text. The tasks could be submitted in English or Spanish, depending upon the language typically used for language arts instruction. The types of texts chosen by the teachers (e.g., fiction or nonfiction) and the design of the task were completely at the discretion of the teachers. To give teachers a sense of the range of the types of tasks that would fall within our guidelines, teachers were told that the tasks submitted may include (but were not limited to) an evaluation of a book, analysis of a character, a summary of a text, or an essay comparing and contrasting texts. Teachers were asked to provide only tasks that they considered to be challenging. The purpose of asking for challenging tasks was to provide insight into what teachers considered rigorous, high-level writing tasks, and, in so doing, establish a common basis for comparison.

For each task, teachers chose four samples of student work—two from students who were considered to produce “high-quality” work and two from students who produced “medium-quality” work; teachers labeled these as “high” or “medium.” Teachers were asked to provide student work in the language of instruction they normally used. All tasks for which students responded to Spanish language texts and wrote in Spanish were analyzed for this study ( $N = 56$  tasks,  $N = 224$  samples of student work).<sup>7</sup> The instructions were distributed in September. Tasks and student writing were collected 6 months later. Teachers completed a two-page coversheet for each task describing the instructional context, directions to students, the text, and the criteria used to grade student work. Teachers were paid \$100 for their participation.

### Measures

**Task quality.** The quality of writing tasks was assessed using the Instructional Quality Assessment (IQA; see App. Tables A1 and A2). This measure was developed by researchers at the University of Pittsburgh and Carnegie Mellon (for a full de-

scription of the IQA, including its research base and technical qualities, see Junker et al., 2006) and includes a set of protocols for investigating the quality of tasks in the content areas of literacy and mathematics. Previous research has demonstrated that the IQA ratings of task quality positively and independently predict student achievement in reading and mathematics (Matsumura et al., 2008). In this study, the IQA was used to measure two dimensions of task quality:

- Quality of the text (1 = low, 3 = high): The degree to which a text contains literary or informational content that is complex and engaging enough to support extended written responses (Beck & McKeown, 2001). This dimension also considers the richness and variety of the language in the text.
- Cognitive demand of the task (1 = low, 4 = high): The degree to which a task supports students to apply complex thinking skills and use appropriate evidence and details from a text (Clare & Aschbacher, 2001; Matsumura, Patthey-Chavez, et al., 2002; Newmann et al., 2001).

A children's literature specialist scored the texts and a member of the research team scored all tasks. Another member of the research team rated a subsample of tasks (12 of 56) on both dimensions. Raters were never off by more than one for either dimension. Exact match agreement for ratings of quality of the text and cognitive demand were 83% and 75%, respectively. Cohen's kappa for each rating was .65, indicating substantial agreement.

**Academic language.** Students' written work was assessed by analyzing several discrete linguistic features that are key indicators of students' facility with academic language, organized into two conceptual categories: lexical and grammatical features. Additionally, we examined total word count and overall quality, a holistic score based on the NAEP rubric. Below, we explain how we operationalized each feature, the distribution of scores for each variable, and the results from interrater reliability analyses. Finally, we describe the statistical analyses employed with each variable given the different distributions of our outcomes.

#### *Lexical features.*

General academic vocabulary—the total number of words that represent a “prestige lexis” (Schleppegrell, 2004). These are words that represent sophisticated, nuanced meanings and appear in a range of academic texts from many disciplines, such as *angustia* (anguish) and *agradecer* (to be grateful for).

Specialized academic vocabulary—the total number of different words that are part of a “prestige” or “specialist lexis” (Schleppegrell, 2004) that convey technical meanings within a given discipline, such as *polinadores* (pollinators) and *istmo* (isthmus).

Intermediate words—the total number of words in the sample that do not appear in the top 500 most frequent words in Spanish texts per the Juilland and Chang-Rodriguez Spanish word frequency list (1964) and are not general or specialized academic vocabulary words. Examples are *esconder* (to hide), *enfermedad* (sickness), *orgullosa* (proud), *vestido* (dress/suit), *subir* (to go up), and *equipo* (team).<sup>8</sup> These words signal the use of a more varied lexis (Colombi, 2002; Laufer & Nation, 1995) or a “stepping stone” toward development of an academic register.

Distribution of scores: Our count variables for all three vocabulary variables reflected a Poisson distribution. All three distributions were negatively skewed. For



general academic vocabulary ( $M = 1.27$ ,  $SD = 1.64$ ), half of the students in the sample included no more than one general academic vocabulary word, and there was a large range (0–11). More than half of the samples did not include any specialized academic vocabulary words ( $M = 0.67$ ,  $SD = 1.24$ ) and ranged from 0 to 8. For intermediate vocabulary ( $M = 13.44$ ,  $SD = 9.28$ ), the distribution ranged from 1 to 43. Our statistical analyses account for these highly skewed distributions by employing hierarchical linear models allowing for a Poisson distribution.

**Interrater reliability:** Two raters came to consensus to categorize all words as general or specialized academic vocabulary or intermediate words, building an inventory of each word type while coding to optimize consistency of ratings. Next, two independent raters examined a stratified random sample of a corpus of 50 Spanish words (20 general academic and 20 specialized academic vocabulary words, plus 10 high-frequency distracter words) from the writing samples and placed them into one of the three categories. This allowed for three paired comparisons. Exact match agreement between the pairs on the 50 words were 80%, 84%, and 88%. Cohen's kappa for each pair of raters ranged from a low of .68 to a high of .81, indicating substantial agreement among the raters.

*Grammatical features.*

**Embedded clauses—**frequency count of clauses that are embedded within the noun, verb, or adverbial group of a higher-ranked clause (e.g., *Los indígenas limpiaba la basura que estaba en el río*. [The Native Americans cleaned up the trash that was in the river.]). Coding decisions adhered to Martin, Matthiessen, & Painter's (1997) guidelines.

**Variety of connectives—**the number of different functions represented by connectives present in the sample. Eight categories were coded: additive, elaborative, temporal, comparative, causal, adversative, continuous, and contrastive. Scores ranged from zero (no connectives present) to eight (all types present). Pertinent categories were distilled from Halliday (1994) and Sanders, Spooren, and Noordman (1992).

**Causal connectives—**the presence of connectives indicating a causal relationship between text ideas, typically marked by *porque* (because) and *ya que* (since) (e.g., *Le estaban echando plástico, basura y cosas para que se contaminara*. [They were throwing plastic, trash, and other things so that it would become polluted.]).

**Temporal connectives—**the presence of connectives indicating a temporal relationship between text ideas or indicating a logical sequence of information marked by connectives such as *cuando* (when), or *primero*, *segundo*, and *finalmente* (first, then, finally) (e.g., *Primero el río estaba limpio y bonito pero después lo empezaron a ensuciar el río*. [First the river was clean and pretty, but afterward they began to pollute the river.]).

**Adversative connectives—**the presence of connectives indicating an adversative relationship between text ideas, typically marked by *pero* (but) and *sin embargo* (although/nonetheless) (e.g., *Esperanza era una niña adinerada al igual que su madre . . . sin embargo, todo ha cambiado para ellas*. [Esperanza was a wealthy girl just like her mother . . . nonetheless everything has changed for them.]).

**Distribution—**the distribution of embedded clauses ( $M = 3.24$ ,  $SD = 3.23$ ) was highly skewed with a Poisson distribution. The values ranged from 0 to 18; three-quarters of the samples contained between 0 and 2 embedded clauses. The distribution of the values for variety of connectives ( $M = 3.32$ ,  $SD = 1.54$ ) was normal,

allowing for statistical modeling of a continuous variable. For causal, temporal, and adversative connectives, the writing samples displayed bimodal distributions. Correspondingly, we created dichotomous variables. In the corpus, 71% of samples contained at least one causal connective (50% contained more than one) and 64% contained at least one temporal connective (46% contained more than one). Given these distributions, we scored causal and temporal connectives 1 if the writing sample contained more than 1, and 0 if the writing sample contained 0 or 1. Only 31% of the writing samples contained at least one adversative connective. Here, the dichotomous variable was 1 for the presence of an adversative connective, 0 otherwise.

*Interrater reliability.* In order to assess the extent to which raters agreed on the creation of these variables, we employed different strategies for measuring interrater agreement. Two raters coded 44 random samples of writing (20% of the total) for each of the variables described above. Since the number of embedded clauses was a count variable, we examined the extent to which raters' counts were correlated. We found a high correlation between raters of .86 (.77 excluding cases where no embedded clauses were present). Examination of the scatterplot did not reveal concern for any outliers in the data. The variable measuring the variety of connectives was bound between 0 and 8. Exact match agreement for this variable was 68%, and agreement ratings were only off by one for 96% of the samples. Cohen's kappa for this variable was .51, indicating moderate agreement between raters. Exact match agreement for the dichotomous variables for causal, temporal, and adversative connectives ranged from 91% to 95%. Cohen's kappa ranged from .76 to .90, indicating substantial agreement between raters.

**Global features.** At the discourse level, academic language is shaped by and reflects conventions for the overall organization and structure of writing. In the present study, students responded to academic tasks in the context of language arts classes. Writing that adheres to genre conventions in this context is organized to share information, communicate interpretations, and provide evidence (Persky, Daane, & Jin, 2003). In addition, length of composition, while itself not an indicator of an academic register, is included here to be consistent with investigations of writing in elementary and secondary schools. Although academic writing tends to be concise and lexically dense (Eggin, 2004; Kress, 1996; Schleppegrell, 2004), and therefore longer compositions are not necessarily higher quality, severe constraints on the length of student essays impair students' ability to communicate complex information and ideas and to provide supporting evidence. Thus word count is important to take into account. The following were analyzed to assess overall quality:

*Overall quality* (1 = low, 6 = high): Six-point rubrics for scoring fourth graders' writing from NAEP accounted for the clarity, coherence, organization, and development of the writing (Persky et al., 2003). A score of 6 indicates that the essay was well organized with strong transition, contained a variety of sentence structures and appropriate word choice, and successfully developed an idea using effective, supporting details.

*Total words:* The total number of words in the student writing sample.

*Distribution:* Overall quality ratings ( $M = 3.18$ ,  $SD = 1.09$ ) were normally distributed. More than 66% of the samples were rated "uneven" (3) or lower; over one-quarter of the pieces were scored "insufficient" (2). We observed a tremendous range in total words ( $M = 120$ ,  $SD = 77$ ), from 18 to 370. The overall distribution was normal.

*Interrater reliability:* For overall quality, 120 random samples (approximately 50%) were coded for reliability. Exact match agreement was 59%, and agreement rates were only off by one for 96% of the samples. The corresponding Cohen's kappa was .41, indicating a moderate level of agreement between raters. It should be noted that after discussing the jointly coded pieces, the raters arrived at a consensus code that was used as the final outcome in analyses. For total words, 45 random samples (20%) were chosen; raters had perfect agreement.

### Statistical Analyses

Multilevel models were used to investigate the degree of variation between tasks on the various outcome measures of student writing. By sampling multiple students per task and multiple tasks per teacher, we nested student essays within a given task, and tasks within teachers. The resulting fully unconditional three-level model produces variance components between students on the same task, between tasks within teachers, and between teachers. Of primary interest in these models is examining how much variance is partitioned between tasks and how much is between teachers. Significant portions of the variance between tasks within teachers would indicate that students' compositions vary in important ways as a function of the tasks assigned to students. Large amounts of variance at the teacher level could indicate either that students are unevenly distributed to classrooms or, alternatively, that teachers significantly vary in the general rigor of their tasks and/or their expectations for student work.

Expanding on the results of the variance decomposition models, we also examined prediction models in which the cognitive demand of the task assignment and the quality of the text used were the main independent variables of interest. We used these two covariates to predict our academic language outcomes. We were interested not only in whether significant variation exists between tasks on the outcomes, but also in whether the presence of academic language covaries with the cognitive demand of the task and the quality of the text assigned.

In addition, to investigate whether variation in students' use of the features of academic language might be a function of other student and classroom characteristics, we also examined prediction models that included proxies for student achievement. At the student level, we used teachers' judgment of whether students' work was of "high" or "medium" quality. (It was not feasible to collect data that identified students when teachers submitted writing samples to us.) Additionally, we included a variable at the teacher level indicating the class average score on the TAKS from the previous year's administration to adjust for differences in prior achievement.<sup>9</sup>

We ran hierarchical linear models on all of the academic language outcomes described previously. Below we describe the three-level model for the dichotomous variables indicating use of different types of connectives—causal, temporal, and adversative. Other statistical models are variants of this analysis utilizing different sampling distributions at level 1 of the model. For dichotomous outcomes, a three-level hierarchical binomial model was used where the input was a dichotomous variable. For example, for the variable adversative connectives, a value of 1 was used if an adversative connective was present, 0 otherwise.<sup>10</sup> The general form for the prediction models was as follows: Level 1 (student essays):

$$\eta_{ijk} = \log[\varphi_{ijk}/1 - \varphi_{ijk}] = \pi_{ojk} + \pi_{ijk}(\text{Hi/Med}). \quad (1)$$

Here,  $\eta_{ijk}$  is the log odds of evidence of the use of adversative connectives,  $\pi_{ojk}$  is the mean outcome for task  $j$  in teacher  $k$ , (Hi/Med) indicates whether the student was considered to produce “high” or “medium” quality work by the teacher, and  $\pi_{ijk}$  is the association between this achievement designation and the outcome for task  $j$  in teacher  $k$ .

Across tasks within teachers, the proportion of essays employing the academic language feature varies as a function of the quality of the task and a random student error. Level 2 (tasks):

$$\begin{aligned} \pi_{ojk} &= \beta_{0ok} + \beta_{01k}(\text{Cog. Dem.}) + \beta_{02k}(\text{Text Quality}) + r_{pj k}, & (2) \\ \pi_{1jk} &= \beta_{1ok}. \end{aligned}$$

Here,  $\beta_{0ok}$  is the average incidence of connectives in teacher  $k$ , Cog. Dem. is the standardized task rating for the cognitive demand of the task,  $\beta_{01k}$  is the effect of cognitive demand of the task on student use of the feature of academic language, Text Quality is the standardized rating for the quality of text, and  $\beta_{02k}$  is the effect of text quality on the student use of the feature of academic language. The random effect,  $r_{pj k}$ , is assumed normally distributed with mean 0 and variance  $\tau_{\pi}$ .

Looking across teachers, we obtained the proportion of all essays where a particular feature of academic language occurred, on average. Level 3 (teachers):

$$\begin{aligned} \beta_{0ok} &= \gamma_{000} + \gamma_{001}(\text{TAKS}) + \gamma_{002}(\text{Miss TAKS}) + u_{0ok}, & (3) \\ \beta_{01k} &= \gamma_{010}, \\ \beta_{02k} &= \gamma_{020}, \\ \beta_{1ok} &= \gamma_{100}. \end{aligned}$$

Here, the intercept  $\beta_{0ok}$  varies randomly between teachers, while the effects of the task characteristics  $\beta_{0qk}$  are fixed. In these models, the coefficient  $\gamma_{000}$  describes the proportion of student essays within teachers with the presence of the academic language outcome, (TAKS) is the standardized class average prior achievement from the Texas state test,  $\gamma_{001}$  is the influence of prior achievement on presence of connectives, (Miss TAKS) is an indicator of whether or not prior achievement data were available, and  $\gamma_{002}$  is the corresponding association between missing data and the presence of connectives. Our foci in these models are  $\gamma_{010}$  and  $\gamma_{020}$ , which are the coefficients for the effects of the cognitive demand of the task and text quality, respectively, on the presence or absence of connectives. Finally,  $u_{0ok}$  is a random error assumed to be normally distributed with variance  $\tau_{00}$ .

We also examined the frequency with which other indicators of academic language were evident in student writing. These included Poisson models, where the outcomes were highly skewed counts for the following variables: the number of general academic, specialized academic, and intermediate vocabulary words, and counts of embedded clauses. These also included continuous models, where the distributions were normal: variety of connectives, overall quality, and total words. These models take the exact same form as the binomial models just described, with the exception of the level 1 sampling distribution. For Poisson models, the output is interpreted in terms of event ratios (sometimes also referred to as incidence rate ratios). For continuous models, it is assumed that the outcome has a normal distribution and that the residuals are normally distributed with a mean of 0 and standard

deviation of  $\sigma^2$ . For all of our analyses, we examined residuals from the hierarchical regressions to assess the extent to which they met basic assumptions of linear regression models; all were satisfactory.

## Results

### To What Degree Do Students Use Salient Features of Academic Language When Writing in Their Native Language (Spanish)?

Means and standard deviations for each feature of academic language as well as for overall quality and total words are presented with variable distributions in the Method section. Two observations should be noted. First, student writing evidenced a very limited lexis. While students included on average between one and two general academic vocabulary words in their writing such as *elegante* (elegant) and *cautelosamente* (cautiously), half of the samples contained no more than one general academic vocabulary word. Students used some specialized academic vocabulary such as *descomponedor* (decomposer) and *atómica* (atomic), but use of such words was very rare; more than half of the samples did not include any specialized academic vocabulary words. While it might not be surprising that most students in the upper elementary grades did not include general or specialized academic vocabulary in their writing, it is notable that for half the students, less than 11% of the words included in their samples represented intermediate words, such as *participar* (to participate), *nieve* (snow), and *divertido* (fun). Thus for half of the students in this sample, nearly 89% of the words they included in their essays were among the top 500 most frequently used words in Spanish; this illustrates the limited range of word choice in the samples. These results suggest that students generally did not select vocabulary that would enable them to convey precise meaning or project an authoritative stance toward their position about a text.

Second, although students employed causal and temporal connectives frequently, they used adversative connectives more rarely. Students employed causal and temporal connectives to explicitly communicate the relationships between text ideas and information. For example, temporal connectives often signaled chronological relationships, as seen in an excerpt from an essay about Mae Jemison: “*Logró ser médico y luego se unió al Cuerpo de Paz. . . Después participó en el entrenamiento de astronautas de la NASA*” (She succeeded in becoming a doctor and next she joined the Peace Corps. . . Afterward she participated in NASA’s astronaut training).

On the other hand, grammatical features that signal adversative relationships and embedding to promote the condensation of language were rare. One example of a student’s use of an adversative connective to orient the reader to the structure of the essay is from a composition about Pam Muñoz Ryan’s *Esperanza renace* (*Esperanza Rising*), in which “*pero sin embargo*” (but nonetheless) is employed by the student to mark a shift from discussion about the characters’ previous wealthy lifestyle to their plunge into poverty: “*Pero, sin embargo, todo ha cambiado para ellas*” (But, nonetheless, everything has changed for them).

Finally, students included an average of just over three embedded clauses per essay, enabling them to condense and package information (usually into noun clauses) and allowing them to connect or foreground important ideas. For example, one student wrote in response to *Esperanza renace* (*Esperanza Rising*): “*Al día sigui-*

Table 1. Quality of Writing Tasks (*n* = 56 Tasks)

	Median	SD	Range	Poor (1) %	Basic (2) %	Good (3) %	Exemplary (4) %
Quality of the text <sup>a</sup>	2	.68	1–3	14.3	44.6	33.8	n/a
Cognitive demand of the task	2	1.9	1–4	32.1	50	16.1	1.8

<sup>a</sup> The percentages do not add up to 100 because four texts (7.1%) could not be located and rated.

*ente Esperanza lucía uno de los vestidos que abuela y Miguel le habían comprado*” (The following day, Esperanza was wearing one of the dresses that her mother and Miguel had bought her). In this excerpt, the fact that the dress was a gift from her (previously destitute) mother is essential information to this student’s analysis of how and why the main character, Esperanza, changed in the novel; this meaning is realized in the student’s grammar by embedding information about the source of the dress into the noun clause.

### What Is the Quality of Writing Tasks Assigned to Native Spanish Speakers?

As shown in Table 1, one-third of the tasks were based on rich extended texts. For fiction, these generally were contemporary novels written for a preadolescent audience, such as *Esperanza renace* (*Esperanza Rising*). Nonfiction texts that received the highest score developed conceptual understanding in a content area using domain-specific vocabulary and often used many features of nonfiction texts (e.g., annotated graphics, table of contents) such as *Nuestra fantástica tierra* (*Our Wonderful Earth*) by Nicola Baxter. Over one-third of the texts were expository (19 of 56 texts). Less than half (44.6%) of the texts students responded to were considered “basic.” Basic texts included short selections from a basal reader; texts with simple plots intended for younger readers, such as *Irene la valiente* (*Brave Irene*) by William Steig; or brief articles (a few paragraphs) from a current events magazine. Just over 14% of texts received the lowest score as they contained only a very simple narrative and simple language, such as *El maíz maravilloso de México* (*Mexico’s Marvelous Corn*) by Margarita González-Jensen. One-third of all tasks (*n* = 76) required students to respond to nonfiction texts; the remaining required students to respond to fiction.

Relatively few writing tasks (17.9%) guided students to apply higher-level skills and write extended responses (i.e., scores of 3 or 4). For example, one high-cognitive-demand writing task had students use examples from the text to explain the significance of the title *Esperanza renace* (*Esperanza Rising*). A cognitively demanding writing task in response to a nonfiction text guided students to explain their understanding of an ecosystem. Students were asked to provide “examples of an ecosystem, name the parts of an ecosystem, show that complex systems may not work if some parts are removed.” Half of the tasks were considered to be of basic quality, meaning that students were asked to write a cursory summary of a text. For example, a typical description of a basic task was, “Students read an internet article about the China earthquake [*sic*] and wrote a summary report” in which they were asked to provide “title, one paragraph, main idea, supporting details.” Finally, nearly one-third of the tasks (32.1%) were considered to be below basic because they guided students to retrieve isolated facts from a text or write on a topic that was only tangentially connected to the text. For example, one task guided stu-

Table 2. Task Quality Ratings Predicting Lexical Features of Academic Language

	General Academic Vocabulary, Coef. (SE)	Specialized Academic Vocabulary, Coef. (SE)	Intermediate Vocabulary, Coef. (SE)
Intercept	.15 (.14)	-.51 (.19)	2.56 (.08)
Classroom level:			
Prior achievement TAKS	.17 (.17)	-.41 (.24)	-.17 (.10)
Missing TAKS	-.16 (.36)	.43 (.50)	-.10 (.21)
Task level:			
Quality of text	.12 (.13)	.40 (.19) *	.08 (.07)
Cognitive demand	.21 (.14)	.43 (.19) *	.27 (.08) **
Student level:			
High/medium	.17 (.13)	.34 (.13) *	.18 (.05) **
Variance components:			
Variance between tasks	.38	1.21	.17
Variance between teachers	.13	.19	.05
% variance explained	15	7	33
Fully unconditional models:			
Variance between tasks (%)	.30 (50)	1.39 (93)	.15 (45)
Variance between teachers (%)	.30 (50)	.12 (7)	.18 (55)
Total variance	.60	1.51	.33

\*  $p < .05$ .\*\*  $p < .01$ .

dents to complete a “nonfiction summary star” in which they were required to transcribe the title into the center of the star and in each of the five points respond to “Who? What? Where? When? and Why?” As mentioned previously, teachers were asked to provide tasks that they considered to be challenging and were given several months to collect writing samples before submitting materials. Thus there is no reason to believe that the quality of the tasks we collected was constrained by a time limit.

#### Does Students’ Use of Academic Language Vary as a Function of Individual Teachers, Tasks, and Students? If So, Does the Quality of Tasks Predict Students’ Use of Salient Features of Academic Language?

In order to learn about the variance decomposition of our indicators of academic language, we examined a series of hierarchical linear models. Tables 2 and 3 present outcomes related to lexical and grammatical features indicative of an academic register. In Table 4 we present the results from global indicators of writing quality. For each outcome, we examined the fully unconditional model before examining a prediction model.

The results portray a consistent theme with respect to the variance decompositions: there was significant variation between tasks within teachers. For example, Table 2 shows that the task was a significant source of variation in the amount of general academic vocabulary students included in their writing. About half of the variance in the fully unconditional model (reported in the bottom of the Table) was found between tasks, while the other half was found between teachers. This indicates that much of the variance in students’ use of the features of academic language could potentially be accounted for by the quality of the task.

Table 3. Task Quality Ratings Predicting Grammatical Features of Academic Language

	Embedded Clauses, Coef. (SE)	Variety of Connectives, Coef. (SE)	Causal, Coef. (SE)	Temporal, Coef. (SE)	Adversative, Coef. (SE)
Intercept	1.08 (.10)	3.35 (.16)	.00 (.24)	-.08 (.19)	-.81 (.21)
Teacher level:					
Prior achievement TAKS	-.15 (.12)	.35 (.20) <sup>+</sup>	-.02 (.31)	.19 (.25)	.26 (.27)
Missing TAKS	.19 (.26)	-.38 (.41)	.16 (.63)	.35 (.50)	-.16 (.55)
Task level:					
Quality of text	-.06 (.08)	.17 (.15)	.14 (.24)	.40 (.20) <sup>*</sup>	.21 (.21)
Cognitive demand	.46 (.09) <sup>***</sup>	.38 (.16) <sup>*</sup>	.74 (.27) <sup>**</sup>	.72 (.22) <sup>*</sup>	.42 (.22) <sup>+</sup>
Student level:					
High/medium	.27 (.09) <sup>**</sup>	.18 (.15)	.46 (.24)	.58 (.27) <sup>*</sup>	-.14 (.27)
Variance components:					
Variance between students	—	1.27	—	—	—
Variance between tasks	.13	.57	1.56	.74	.80
Variance between teachers	.12	.16	.09	.07	.21
% variance explained	58	15	18	33	35
Fully unconditional models:					
Variance between students (%)	—	1.26 (54)	—	—	—
Variance between tasks (%)	.16 (26)	.47 (20)	1.57 (78)	1.00 (83)	.85 (54)
Variance between teachers (%)	.46 (74)	.61 (26)	.45 (22)	.20 (17)	.71 (46)
Total variance	.60	2.34	2.02	1.20	1.56

<sup>+</sup> *p* < .10.  
<sup>\*</sup> *p* < .05.  
<sup>\*\*</sup> *p* < .01.  
<sup>\*\*\*</sup> *p* < .001.

The other outcomes investigating the presence of features of academic language in students' writing also appeared to vary more by task than by teacher. In particular, the presence of temporal connectives, causal connectives, and specialized academic vocabulary all varied largely by task (with 83%, 78%, and 92% of the variance between tasks). One way of interpreting these results is that students are likely to use these features when the task itself explicitly or implicitly calls for them. For many of the other dimensions of academic language—presence of adversative connectives, generalized academic vocabulary, and intermediate vocabulary—about half of the variance was between tasks and about half was between teachers. The use of embedded clauses varied more between teachers than between tasks, suggesting that embedding of clauses, regardless of the task, was used by students in some classrooms more than others (perhaps indicating that they were modeled more in some classrooms than others). Finally, the variety of connectives varied largely as a function of the student, with relatively equal portions divided between the task and the teacher.

The above analyses demonstrate that the majority of features of academic language in writing vary considerably by the tasks presented. In contrast, the overall quality of the writing, which to some degree reflects the effectiveness of an academic voice globally, accentuates the differences between teachers. Table 4 presents the results of the fully unconditional models for the outcomes examining total words and the overall quality of students' writing. Take, for example, the results of the total-words variable. The variance components reveal that 35% of the variance is between students on the same task,<sup>11</sup> 20% is between tasks, and 45% is between



Table 4. Task Quality Ratings Predicting Global Measures of High-Quality Writing

	Total Word Count, Coef. (SE)	Overall Quality, Coef. (SE)
Intercept	123.14 (9.64)	3.18 (.14)
Teacher level:		
Prior achievement TAKS	-.15 (11.34)	-.01 (.17)
Missing TAKS	-12.85 (25.62)	.07 (.38)
Task level:		
Quality of text	4.80 (7.19)	.17 (.10)
Cognitive demand	33.43 (7.94)***	.36 (.11)**
Student level:		
High/medium	21.70 (5.80)***	.33 (.09)***
Variance components:		
Variance between students	1,836.61	.46
Variance between tasks	1,024.28	.17
Variance between teachers	1,496.64	.34
% variance explained	27	22
Fully unconditional models:		
Variance between students (%)	2,098.87 (35)	.50 (40)
Variance between tasks (%)	1,182.84 (20)	.18 (14)
Variance between teachers (%)	2,704.11 (45)	.57 (46)
Total variance	5,985.82	1.25

\*\*  $p < .01$ .\*\*\*  $p < .001$ .

teachers. For global ratings, the majority of the variance in student writing lies between teachers. Nevertheless, given that a significant portion of variance exists between tasks within teachers, we further examined whether it would be possible to predict this variation with the addition of task ratings to the models.

Indeed, the prediction models in Tables 2–4 reveal that the ratings of the cognitive demand of the task are a robust predictor across a number of academic language outcomes. For seven of the eight outcomes examined in Tables 2 and 3, the higher the cognitive demand of the task, the more likely it was that certain features of academic language were present in student writing. This was true even after adjusting for prior student achievement. Here, it was interesting to note that class prior achievement (the class average scale score on TAKS) had little influence on the outcomes, but whether the students were high or middle achieving (i.e., whether the teacher identified the student as one who typically produced high or medium quality work) often did. The inclusion of prior achievement with the task ratings often explained a good portion of the variance between tasks and between teachers. In most cases, the variables explained 15%–35% of the variance, primarily that between teachers. We find it noteworthy that other teacher characteristics (e.g., experience, certification) were not related to the student writing outcomes.<sup>12</sup>

Moreover, the cognitive demand of the task was a positive predictor for both overall quality and total words. Tasks 1 standard deviation higher for cognitive demand resulted in an additional 33 words on the student essay, or about a 27% increase over the average (122 words). This prediction model explained 27% of the variance in the outcome—including a large portion of the variance between teachers. Similar results were found for overall quality, where 1 standard deviation higher in cognitive demand was associated with a higher score by about .36 points (or an additional 11%). This model explained 22% of the total variance in student scores.

The measure of text quality was less successful at predicting use of academic language. However, the coefficient associated with the rating of text quality was nearly always positive, and the ratings were significant for two of the 10 outcomes (specialized academic vocabulary and temporal connectives). Thus, students responding to higher-quality texts appear more likely to incorporate some features of academic language in their written responses. Combined with the ratings of task cognitive demand, these findings suggest an association between tasks and the types of student writing outcomes examined here.

### Examples of High- and Low-Cognitive-Demand Tasks and Students' Use of Academic Language

To provide insight into our linguistic coding scheme and a window into the types of writing tasks teachers submitted as examples of "challenging" assignments, below we present examples of high- and low-quality tasks and corresponding examples of students' written work.

**High-quality writing task.** For this task, which received the highest score for the quality of the text and cognitive demand of the task, grade 4 students read the chapter book *Esperanza renace* (*Esperanza Rising*), a narrative that addresses themes of social class, identity, and renewal. This story is about a very wealthy family in Mexico who, after the father's death, are forced to flee to the United States and work in a farm labor camp. Students were guided to write a six-paragraph essay analyzing how two characters in the book changed from the beginning of the story up to the present chapter.

Figure 1 presents an essay considered by the teacher to be of high quality for the class. This student's response to the task included more features of academic language than was typical in the corpus. For overall quality, the student writing received the highest score (6) due to the writer's effort to introduce the reader to the characters and provide a rudimentary but clear orientation to the conflict faced by the characters in this novel. The student then claims that two of the characters changed in parallel over time and provides extensive evidence from the text, noting similarities between the characters. The essay closes with a brief statement evaluating the quality and organization of the novel.

This piece demonstrates early and developing efforts to present an analysis and provide evidence from the text to back up the student's claims. First, use of embedded clauses was more prevalent in this response (containing six embedded clauses) than was observed in the samples on average (three embedded clauses per writing sample). For example, when the student wrote, "Uno de los problemas más pesados para ellas fue *cuando el tío Luís quema su casa y hace que pierdan todo lo que tenían*" (One of the most difficult problems for them was when Uncle Luis burned their house and made them lose everything that they had; embedded clause complex underscored), the student successfully presents a major event in the novel and its outcomes as "things." By doing so, the student conveys the assertion with authority, as if stating a fact.

Second, the student conveys temporal, causal, and additive relationships between text ideas by using connectives such as *in the beginning*, *since*, *also*, and *and*. By employing these connectives, the student orients the reader to the chronological sequencing of the composition, provides some motivations for the characters' actions, and offers elaboration through expansion of the points she presents. While the additive connectives employed are not sophisticated (*and*, *also*), the student uses

Spanish Transcription	English Translation
<p>En el cuento <i>Esperanza Renace</i> hay dos <u>personajes principales</u>, la Sra. Ramona (mamá) y Esperanza. Ellas tienen que pasar por muchos problemas <b>después de que</b> el Sr. Ortega muere. Los problemas habían empezado en Aguascalientes, México. Uno de los problemas más pesados para ellas fue <i>cuando el tío Luis quema su casa y hace que pierdan todo lo que tenían</i>. <b>Así que</b> ellas tienen que <u>escapar ya que</u> si se quedaban la Sra. Ramona se tendría que casar con el tío Luis y se tendría que separar de Esperanza. <b>Entonces</b> las dos deben de hacer grandes cambios en su vida.</p> <p>La Sra. Ramona <b>al principio</b> del cuento era muy <u>adinerada</u>, se recogía el pelo muy <u>elegantemente</u>. Ella se vestía con muchos vestidos elegantes. No tenía que trabajar, hacer que hacer <b>ya que</b> tenía <u>empleados que hacían todo por ella</u>. <b>También</b> ella era la que <i>daba las órdenes a los empleados</i>.</p> <p>Esperanza era una niña adinerada al igual que su madre. Tenía ropa de muy buena calidad. Como sus vestidos de <u>seda</u>. <b>También</b> tenía una gran recámara, su cama grande y tenía muchas muñecas <i>que su padre le había regalado</i>. Ella tenía todo <i>lo que quería</i>. No tenía nada <i>de qué hacer ya que</i> los empleados también hacían todo por ella.</p>	<p>In the story <i>Esperanza Rising</i> there are two <u>main characters</u>, Mrs. Ramona (mom) and Esperanza. They have to deal with many problems <b>after</b> Mr. Ortega dies. The problems had started in Aguascalientes, Mexico. One of the most difficult problems for them was <i>when Uncle Luis burns their house and made them lose everything that they had</i>. <b>So</b> they had to <u>escape since</u> if they stayed Mrs. Ramona would have to marry Uncle Luis <b>and</b> she would have to separate from Esperanza. <b>So</b> both of them had to make big changes in their lives.</p> <p><b>In the beginning</b>, Mrs. Ramona was very <u>wealthy</u>, she fixed her hair very <u>elegantly</u>. She dressed in elegant clothes. She did not have to work, she did what she wanted to do <b>since</b> she had <u>employees that did everything for her</u>. <b>Also</b> it was she <i>who gave orders to the employees</i>.</p> <p>Esperanza was a wealthy girl just like her mother. She had well made clothes. Like her <u>silk</u> dresses. <b>Also</b> she had a great bedroom, a large bed, <b>and</b> she had many dolls <i>that her father had given to her</i>. She had everything <i>that she wanted</i>. She didn't have anything <i>she had to do since</i> the employees would do everything for her.</p>

Figure 1. Example of student writing in response to a high-quality task. Embedded clauses are italicized, connectives (additive, elaborative, temporal, comparative, causal, adversative, continuous and contrastive) are in bold, and general and specialized academic vocabulary is underlined. Note that because our research focused squarely on the presence of academic language features in student writing and we were not interested in students' proficiency with Spanish orthography, transcriptions do not reflect students' spelling errors.

<p><b>Pero, sin embargo</b>, todo ha cambiado para ellas. La Sra. Ramona ha cambiado mucho después de que el Sr. Ortega muere. Ella se enoja mucho con el tío Luis le dice que sí se caza con él. <b>También</b> porque él quiere ser el que manda. <b>Pero</b> como la Sra. Ramona <u>rechaza su propuesta</u> el tío Luis enciende su casa. Ahora ya no tenía nada, todo se había hecho cenizas. <b>Así que</b> ellos escapan para California.</p> <p>Lo mismo había pasado con Esperanza. De niña adinerada, ahora ya no tenía nada. <b>Ahora</b> ya no era tan feliz como antes, <b>ahora</b> estaba triste <b>ya que</b> todo lo había perdido. Tendría que usar ropa <i>de otra gente había donado</i>. Eso había sido un gran cambio para ella, <b>ya que</b> estaba <u>acostumbrada</u> a usar buena y bonita ropa.</p> <p>Yo opino que este cuento es muy bueno. Pienso que está muy bien <u>organizado</u>. Empieza feliz, <b>luego</b> se pone triste. Los eventos para mí tienen mucho sentido <u>conforme</u> va la historia.</p>	<p><b>But nonetheless</b>, everything has changed for them. Mrs. Ramona has changed a lot after Mr. Ortega dies. She becomes angry a lot with Uncle Luis [who] tells her that she should marry him. <b>Also</b>, because he wants to be the one in charge. <b>But since</b> Mrs. Ramona <u>rejects his proposal</u>, Uncle Luis burns their house. Now they had nothing, everything had burned to ashes. <b>So</b> they escaped for California.</p> <p>The same thing happened with Esperanza. From being a <u>wealthy child</u>, she now had nothing. <b>Now</b> she wasn't happy like before, <b>now</b> she was sad <b>since</b> everything had been lost. She would have to wear used clothing <i>that other people had donated</i>. This had been a great change for her <b>since</b> she was <u>accustomed</u> to wearing good and pretty clothes.</p> <p>My opinion is that this story is very good. I think that it is well <u>organized</u>. It begins happy, <b>then later</b> it becomes sad. The events make a lot of sense to me <u>according to the</u> rest of the story.</p>
--	--

Figure 1 (continued).

them to provide additional evidence that is later used to support claims about how the characters change over time. This essay was longer than was typical of the corpus, containing 368 words in comparison to the average of 121 words.

Evidence of lexical features of academic language was also relatively abundant in the student's response to this rigorous task. First, the student included general aca-

demical vocabulary, such as *adinerada* (wealthy) and *elegantemente* (elegantly). This student included 10 general academic vocabulary words, nearly eight times more than the average writing sample in the corpus. Second, the sample contained two examples of specialized academic vocabulary (e.g., *personajes principale* [main characters]). While this might be considered nominal, it is important to consider that half the samples in the overall corpus contained no specialized academic vocabulary. Finally, 43% of words used in this sample were classified as intermediate vocabulary; that the proportion of words in the sample that were not high frequency was much higher in this sample than the corpus mean of 13.44% indicates a higher degree of lexical diversity than was typical in the corpus.

**Low-quality writing task.** For this task, given a low score for the quality of the text and the cognitive demand of the task, grade 5 students read two one-page passages about the life of Thomas Alva Edison. Students then were asked to compare and contrast the passages by completing a worksheet about how the texts were similar and different. While the idea of contrasting two nonfiction texts has strong potential for teaching students to consider the sources of informational texts and how the author's perspective shapes these sources, in this case students were guided to focus on discrete biographical facts and were required to present only the barest comparison of the texts. Figure 2 displays an example of student work considered by the teacher to be of high quality for the class.

In contrast to the previous responses and in comparison to the rest of the corpus, the student's response to the task illustrates a near absence of features of academic language. For overall quality, the low score (2) was given because the response would be nearly impossible to understand for a reader who had not read the passages that inform it. In fact, it appears that the student did not receive sufficient support to understand that both texts were about the same historical figure. The relationship between discrete pieces of information is not clear in the response. With respect to total words, the format of the task provided very limited space for students to write their answers. As a result, the student's writing contains 66 words, in contrast to an average of 120 words in the overall corpus. Notably, of all examples of student work submitted in response to this task, this example, deemed high quality by the teacher, was the longest.

There is almost no evidence of the lexical and grammatical features of academic language in this example. To the contrary, several rudimentary grammatical conventions were violated (e.g., sentence fragments such as "*No nos dicen cuando*" [They don't tell us when] and lack of subject-adjective agreement in "*mucho laboratorios*"). The response contained only one embedded clause, reflecting the near absence of grammatical structures that enable writers to condense language. Moreover, the student employed just one additive connective (*y* [and]), contributing to the lack of coherence between discrete pieces of information presented. Finally, the response contained just one general academic vocabulary word (*laboratorios* [laboratories]) and one specialized academic vocabulary word (*telégrafo* [telegraph]). Just 4% of the words in the sample were classified as intermediate, demonstrating that not only did the piece lack academic vocabulary, but in fact nearly all word choice was drawn from the highest-frequency words in Spanish.

These observations of extremely basic and nonspecific word choice, incomplete sentences, and constructions that violate rudimentary grammatical conventions were not isolated to this example but instead were noted across each piece of student

Spanish transcription	English translation
<p>Libro #1:</p> <p><i>Edison Thomas Alva</i></p> <p>No nos dicen cuando.</p> <p>Trabajo de grande.</p> <p>Edison se miraba como una niña.</p> <p>Hizo mucho <u>laboratorios</u>.</p> <p>Se cambia de partes de New Jersey.</p>	<p>Book #1:</p> <p><i>Edison Thomas Alva</i></p> <p>They don't tell us when.</p> <p>He worked when he was older.</p> <p>Edison looked like a girl.</p> <p>He made many <u>laboratories</u>.</p> <p>He moved to a different part of New Jersey.</p>
<p>Libro #2:</p> <p><i>Thomas Edison</i></p> <p>Thomas trabajó muy chiquito a los 12 años. Tenía 3,400 cuadernos <i>llenos de sus ideas de Thomas</i>. Thomas a los 18 de octubre del 1931 trabajó por mucho tiempo. A los 15 años Thomas estaba trabajando en oficina de <u>telégrafos</u>.</p> <p>Lo que tienen los libros en común:</p> <p>Nacieron en la misma vez <b>y</b> murieron de la misma vez. Son los dos científicos.</p>	<p>Book #2:</p> <p><i>Thomas Edison</i></p> <p>Thomas worked very young at 12 years old. He had 3,400 notebooks <i>filled with his ideas</i>. On October 18, 1931, Thomas worked for a long time. At 15, Thomas was working in a <u>telegraph</u> office.</p> <p>What the books have in common:</p> <p>They were born at the same time <b>and</b> they died at the same time. Both are scientists.</p>

Figure 2. Example of student writing in response to a low-quality task. Embedded clauses are italicized, connectives are in bold, and general and specialized academic vocabulary is underlined.

writing in response to this poor-quality task. It is essential to keep in mind that the teacher submitted this example as a highly challenging task. This example of a low-cognitive-demand task is not an outlier, but instead is representative of 32% of tasks in the corpus.

### Discussion

Structural inequities in the education of minority-language students have been well documented (Gándara, Rumberger, Maxwell-Jolly, & Callahan, 2003; Gutiérrez, 2005). Teachers of ELL students have higher mobility rates (Katz, 1999), schools that serve large numbers of ELL students have poorer infrastructures (Gándara et al., 2003), and ELL students are disproportionately represented in nonacademic or low-track classes (Gándara et al., 2003; Harklau, 1994). Much less is understood, however,

about the quality of instruction provided to ELLs in U.S. schools, and the specific aspects of instruction that contribute toward improved academic outcomes. Our findings may contribute toward opening this “black box.”

Specifically, our results indicate that the quality of writing tasks assigned to this population of native Spanish speakers is generally of low cognitive demand. In fact, over four-fifths of the writing tasks were either “poor” or “basic”; that is, they presented, at best, minimal challenge, such as recalling fragmented information or constructing a surface-level summary about a text, despite the fact that many of these tasks were based on rich texts. This finding is commensurate with other research indicating that cognitively demanding tasks are rare, especially in urban schools that serve high numbers of minority students from low-income families and ELL students (Clare & Aschbacher, 2001; Matsumura, Garnier, et al., 2002; Matsumura et al., 2008; Newmann et al., 2001).

Overall, our results indicate that students rarely utilized features of academic language when writing in their native language. Colombi and Roca have argued that, “in our educational system, many Latinos fail to continue the development of advanced literacy in Spanish as they become proficient in English” despite the fact that academic writing “requires a greater mastery of a more extensive range of linguistic features than does everyday, ordinary Spanish” (2003, p. 9). Our findings support Colombi’s (2002) and Colombi and Roca’s (2003) claim and calls into question the assumption that students who receive language arts instruction in Spanish will be able to transfer their command of academic language in Spanish to their writing in English. In fact, results from this study demonstrate that it would be premature to expect evidence of transfer given students’ limited command of an academic register in their native language (Acevedo, 2003).

Finally, and most importantly, we investigated the relationship between characteristics of writing tasks and students’ use of academic language in their writing. We employed hierarchical linear modeling to disentangle the amount of variation explained by the challenge of the tasks from other potential sources of variation (i.e., the teacher and students). We then entered predictors at the task level to determine which aspects of tasks influence students’ tendency to include features of academic language in their writing. Our results indicate that the cognitive demand of tasks was a significant and strong predictor of students’ use of most features of academic language, including academic vocabulary, embedded clauses, temporal and causal connectives, and use of a variety of connectives. Cognitive demand of the task also predicted the overall quality of students’ writing. The quality of the text on which the task was based was not as strong a predictor of students’ use of academic language, holding a significant relationship only with presence of embedded clauses and temporal connectives. It is notable that the relationships between task quality and students’ use of academic language were detected within a relatively small sample of teachers and tasks, suggesting that there could be a strong relationship between these variables.

The participants in this study represent the largest group of ELLs in U.S. schools: native Spanish speakers from low-income backgrounds (Goldenberg, 2008). It is possible that the findings presented here might extend to other students in the United States who similarly are native Spanish speakers schooled in bilingual contexts, attending urban schools in low-income communities. It is also possible that these findings might extend more broadly to monolingual English-speaking stu-

dents. If so, this would indicate that instruction in upper elementary and middle school is generally lacking in pedagogy that supports the development of academic language.

There is, of course, a need to extend this investigation with larger sample sizes. Recent work sheds light on the potential limitations of our findings given that our estimates were derived from a fairly small sample of tasks per teacher (see, e.g., Bell-Ellison, Ferron, & Kromrey, 2008; Clarke, 2008). Researchers employing simulation studies have found that the coefficients and standard errors produced by HLM are fairly robust to the sparse-data problem even in extreme situations (where the average number of cases at a level is two, and also even when singletons make up a fairly high proportion of cases). However, variance components at the group level may be biased upward and, likewise, the standard errors of the variance components, in particular, are likely to be biased upward, making it more difficult to detect significant group-level variation (Clarke, 2008). In light of this, we find it noteworthy that the chi-square test of significance for variation between tasks within teachers was significant for all of our academic language outcomes. While larger samples might provide clearer guidance about the proportion of variance existing at the teacher and task levels in our models, our preliminary analysis suggests that it is, at least, significant and merits further study.

Simulation studies suggest that our fixed effects are not likely to be biased (Clarke, 2008). Furthermore, our analyses explained a fair amount of between-task and between-teacher variance even when cognitive demand and text quality were the only predictors in our models. This provides some evidence for the predictive validity of task-quality ratings. By comparison, other teacher characteristics (holding a master's degree, years of teaching, self-report of writing instruction on an annual survey) failed to predict differences in student writing outcomes.

### Limitations and Future Research

By honing in on one critical factor that shapes ELL students' opportunities to develop academic writing skills—the quality of writing tasks—this study did not account for the many interrelated factors that also influence opportunity to learn. For example, patterns of interaction in classrooms influence students' opportunity to learn how to write for academic purposes (Dyson & Freedman, 2003). While we believe there is value in thoroughly investigating one facet of instruction, clearly students' learning occurs within rich contexts that interact with but also go far beyond task quality.

Our operationalized definition of academic language also could be considered a limitation of this study. Our approach represents a major departure from the SFL method in which writing is studied for the co-occurrence of lexico-grammatical features within the context of the specific purposes and content of particular writing samples to understand how the “constellation” of features function together to accomplish specific purposes (e.g., Schleppegrell, 2004). While that approach enables rich, contextualized analysis of a small number of samples, the goal of this study was to identify trends in use of the lexical and grammatical features that promote an academic register in relation to the rigor of writing tasks across a relatively large and varied corpus. To this end, we analyzed discrete linguistic variables that are strong indicators of academic language. Doing so allowed us to analyze both the patterns of



language choices across a large corpus and the influence of task quality on these specific choices. As illustrated in our examples of high- and low-quality samples of student writing, the discrete features analyzed in this study appeared to be adequate indicators of an academic register. More research is needed, however, to look more closely at the relationship between these indicators and the academic register and content of students' writing.

Also, we were unable to collect data about individual students' gender, number of years in U.S. schools, and their Spanish and English literacy skills, nor did we collect information about variation in the use of Spanish and English by classroom. Such information would explain additional variance (and reduce error) and would provide clearer insight into how task quality relates to students' use of academic language in Spanish.

Another potential limitation of this research was that, in rating the quality of the text to which students responded in their tasks, we did not explicitly assess the degree to which these texts reflected an academic register. Thus our finding that text quality was not a significant predictor of students' use of most features of academic language may not capture the full story. In future work it will be important to examine texts for the same features that were assessed in student writing for a more valid investigation into whether text quality provides opportunities for students to appropriate an academic voice.

Finally, the cross-sectional design of our study prohibited us from investigating how the quality of writing tasks influences students' development of academic language over time. Further research is needed to understand how the learning opportunities and constraints afforded in tasks might influence the development of an academic voice in students writing for both ELL and non-ELL populations.

## Conclusions

Our results suggest that further investigation of the relationship between tasks and students' academic language use is a productive area for further research that could have important implications for teachers' professional development and curriculum design. The fact that students were provided few opportunities to engage with rich tasks and use an academic register in their writing is of grave concern given that academic writing is a gateway skill for success in school. Students who do not develop academic writing skills likely will not be successful in secondary school and certainly will not be able to complete university-level coursework. Providing rigorous writing tasks based on high-quality texts—in combination with explicit instruction about how lexical and grammatical choices function in different social contexts (Christie, 2002)—will likely accelerate and enrich the growth of elementary grade students' repertoires of academic language.

## Appendix A

Table A1. Instructional Quality Assessment Writing Task Rubrics: Quality of the Text

Quality of the Text	Description
3	The text contains lots of “grist” for students to grapple with in a group discussion or in writing. This grist is seen in the complexity of the content (theme, relationships between characters, etc.) and in the writer’s craft (literary language, rich vocabulary, organizational structures).
2	The text contains some “grist” for students to grapple with during group discussion or in writing. There may be some degree of complexity in the content (theme, relationships between characters, etc.) and in the writer’s craft (literary language, rich vocabulary, organizational structures).
1	There is minimal “grist” for students to discuss to make meaning of the story. It may contain a very simple narrative or very basic information, but these are so straightforward that there is nothing about the text that requires extended discussion or writing. For example, the text may be a simplified version of a complex text, or a short excerpt from a workbook.

Table A2. Instructional Quality Assessment Writing Task Rubrics: Cognitive Demand of the Task

Cognitive Demand of the Task	Description
4	The task guides students to engage with the underlying meanings or nuances of a text. Students interpret or analyze a text AND use extensive and detailed evidence from the text to support their ideas or opinions. AND the task provides students with an opportunity to fully develop their thinking (e.g., challenging questions, extended responses, and analytical and interpretive responses).
3	The task guides students to engage with some underlying meanings or nuances of a text. Students may interpret or analyze a text, BUT they use limited evidence from the text to support their ideas or opinions. There is some opportunity for students to develop their thinking (e.g., challenging questions but structured responses).
2	The task guides students to construct a literal summary of the text based on straightforward (surface-level) information OR engage with surface-level information about the text only. The assignment task guides students to use little or no evidence from the text to support their ideas or opinions.
1	The task guides students to recall isolated, straightforward (surface-level) facts about a text OR write on a topic that does not directly reference information from the text. OR, the task guides students in recalling fragmented information about the text.

## Notes

We thank Mariana Achugar, Amanda Godley, and Catherine Snow for comments on earlier versions of this manuscript. In addition, we are grateful to Jeanette Mancilla-Martinez, Andrea Rolla de San Francisco, and Therese Tardio for their collaboration on the linguistic coding of the writing samples. Finally, we would like to thank the *ESJ* editors and reviewers for their comments and suggestions.

1. The TESOL standards were recently revised to recommend use of students’ “first languages and cultures as the foundation for developing academic language proficiency” (TESOL, 2009).

2. Although the use and prominence of certain linguistic features can vary by discipline (e.g., Achugar & Schleppegrell, 2005; Schleppegrell et al., 2004), many of the linguistic features that enable communication of complex information and abstract ideas are common across a range of academic texts in multiple disciplines.

3. These three facets of an academic register are referred to as *field*, *mode*, and *tenor* in SFL.
4. Confirming the importance of this observation, Christie (2002) noted that use of a range of connectives to establish conjunctive relations was a relatively early accomplishment observed by the upper elementary grades in comparison to the writing of younger students.
5. Schleppegrell and Achugar have argued that causal connectives are not typical of academic writing but instead are more commonly used to convey logical relations in spoken interaction (Achugar & Schleppegrell, 2005; Schleppegrell, 2004). Indeed, published academic texts and texts by advanced writers are less likely to include connectives to establish causal relationships. However, causal connectives may play an important role in students' developing capacity to establish logical relations in their writing.
6. It should be noted that while the data were collected over 2 years, teachers only participated in one of those years. Therefore, the teachers are all independent of one another, and the four assignments from each teacher were collected in a 6-month window over a single year of instruction.
7. Several submissions did not include the requisite four samples of student work per task and thus were excluded from the study.
8. Some everyday words in Spanish are cognates of general academic vocabulary in English. Therefore, knowledge of a range of high-frequency words in Spanish also holds the potential to bolster academic vocabulary growth in English for Spanish-English bilinguals.
9. In three classrooms, achievement data were not available. In order to retain those classrooms in the analyses, we imputed a constant value and included a dummy variable in the models indicating whether the classroom was missing achievement data or not.
10. Note that the distributions of the binomial (dichotomous) and Poisson (count) variables were sometimes underdispersed or overdispersed. In all cases we included adjustments within the HLM software to account for under- and overdispersion in the sampling distribution.
11. Note that this portion of variance, sigma squared, contained both variance between students on the outcome measures and also contained measurement error. Thus, any measurement error due to the confound of time or of different students' work being chosen on different assignments was also included in this estimate, so this portion of variance was not only variance that existed between students.
12. We have not included the models with these predictors in our tables for model parsimony. None of these characteristics was a significant predictor in the model; thus it is not surprising that model diagnostics suggested that omitting these variables was preferable. Interested readers can consult the authors for tables with these results.

## References

- Acevedo, R. (2003). Navegando a través del registro formal: Curso para hispanohablantes bilingües. In A. Roca & M. C. Colombi (Eds.), *Mi lengua: Spanish as a heritage language in the United States; research and practice* (pp. 257–268). Washington, DC: Georgetown University Press.
- Achugar, M. (2003). Academic registers in Spanish in the United States. In A. Roca & M. C. Colombi (Eds.), *Mi lengua: Spanish as a heritage language in the United States; research and practice* (pp. 213–234). Washington, DC: Georgetown University Press.
- Achugar, M., & Schleppegrell, M. J. (2005). Beyond connectors: The construction of cause in history textbooks. *Linguistics and Education*, *16*, 298–318.
- Bailey, A. L. (Ed.). (2007). *Language demands of school: Putting academic English to the test*. New Haven, CT: Yale University Press.
- Bailey, A. L., Huang, B. H., Shin, H. W., Farnsworth, T., & Butler, F. A. (2007). *Developing academic English language proficiency prototypes for 5th grade reading: Psychometric and linguistic profiles of tasks; an extended executive summary* (CSE Technical Report No. 720). Los Angeles: UCLA CRESST.
- Beck, I. L., & McKeown, M. G. (2001). Text talk: Capturing the benefits of read-aloud experiences for young children. *Reading Teacher*, *55*, 10–20.
- Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford.

- Bell-Ellison, B. A., Ferron, J. M., & Kromrey, J. D. (2008, August). *Cluster size in multilevel models: The impact of sparse data structures on point and interval estimates in two-level models*. Paper presented at the Joint Statistical Meetings of the American Statistical Association, Denver.
- Carbonaro, W. J., & Gamoran, A. (2002). The production of achievement inequality in high school English. *American Educational Research Journal*, *39*, 801–827. doi:10.3102/00028312039004801
- Christie, F. (2002). The development of abstraction in adolescence in subject English. In M. J. Schleppegrell & M. C. Colombi (Eds.), *Developing advanced literacy in first and second languages: Meaning with power* (pp. 45–66). Mahwah, NJ: Erlbaum.
- Clare, L., & Aschbacher, P. (2001). Exploring the technical quality of using assignments and student work as indicators of classroom practice. *Educational Assessment*, *7*(1), 39–59.
- Clarke, P. (2008). When can group-level clustering be ignored? Multilevel models versus single-level models with sparse data. *Journal of Epidemiological Community Health*, *62*, 752–758. doi:10.1136/jech.2007.060798
- Colombi, M. C. (2002). Academic language development in latino students' writing in Spanish. In M. J. Schleppegrell & M. C. Colombi (Eds.), *Developing advanced literacy in first and second languages: Meaning with power* (pp. 67–86). Mahwah, NJ: Erlbaum.
- Colombi, M. C., & Roca, A. (2003). Insights from research and practice in Spanish as a heritage language. In A. Roca & M. C. Colombi (Eds.), *Mi lengua: Spanish as a heritage language in the United States; research and practice* (pp. 1–24). Washington, DC: Georgetown University Press.
- Colombi, M. C., & Schleppegrell, M. J. (2002). Theory and practice in the development of advanced literacy. In M. J. Schleppegrell & M. C. Colombi (Eds.), *Developing advanced literacy in first and second languages: Meaning with power* (pp. 1–19). Mahwah, NJ: Erlbaum.
- Crowhurst, M. (1990). Teaching and learning the writing of persuasive/argumentative discourse. *Canadian Journal of Education*, *15*, 348–359.
- Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, *49*, 222–251.
- Dressler, C., & Kamil, M. L. (2006). First- and second-language literacy. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language Minority Children and Youth* (pp. 197–238). Mahwah, NJ: Erlbaum.
- Dyson, A. H., & Freedman, S. W. (2003). Writing. In J. Flood, D. Lapp, J. R. Squire, & J. M. Jensen (Eds.), *Handbook of research on teaching the English language arts* (2nd ed.) (pp. 967–992). Mahwah, NJ: Erlbaum.
- Eggs, S. (2004). *An introduction to systemic functional linguistics* (2nd ed.). New York: Continuum.
- Fang, Z., Schleppegrell, M. J., & Cox, B. (2006). Understanding the language demands of schooling: Nouns in academic registers. *Journal of Literacy Research*, *38*, 247–273. doi:10.1207/s15548430jlr3803\_1
- Gándara, P., Rumberger, R., Maxwell-Jolly, J., & Callahan, R. (2003). English learners in California schools: Unequal resources, unequal outcomes. *Education Policy Analysis Archives*, *11*. Retrieved from <http://epaa.asu.edu/epaa/v11n36/>
- Genesee, F., Lindholm-Leary, K., Saunders, W., & Christian, D. (2005). English language learners in U.S. schools: An overview of research findings. *Journal of Education for Students Placed at Risk*, *10*, 363–386. doi:10.1207/s15327671espr1004\_2
- Gibbons, J. (1999). Register aspects of literacy in Spanish. *Written Language and Literacy*, *2*, 63–88.
- Gibbons, J., & Lascar, E. (1998). Operationalising academic language proficiency in bilingualism research. *Journal of Multilingual and Multicultural Development*, *19*, 40–50.
- Goldenberg, C. (2008). Teaching English language learners: What the research does—and does not—say. *American Educator*, *32*, 8–19.
- Graham, S., & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools*. New York: Carnegie Corp.
- Gutiérrez, K. (2005). The persistence of inequality. In J. Flood & P. L. Anderson (Eds.), *Literacy development of students in urban schools* (pp. 288–304). Newark, DE: International Reading Association.
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London: Arnold.
- Harklau, L. (1994). Tracking and linguistic minority students: Consequences of ability grouping for second language learners. *Linguistics and Education*, *6*, 217–244.

- Juilland, A., & Chang-Rodriguez, E. (1964). *Frequency dictionary of Spanish words*. London: Mouton.
- Junker, B. W., Wiesberg, Y., Matsumura, L. C., Crosson, A., Wolf, M. K., Levison, A., & Resnick, L. B. (2006). Overview of the Instructional Quality Assessment. (CSE Technical Report No. 671). Los Angeles: University of California, CRESST.
- Katz, S. R. (1999). Teaching in tensions: Latino immigrant youth, their teachers, and the structures of schooling. *Teachers College Record*, **100**, 809–840.
- Kress, G. (1996). Writing and learning to write. In D. Olson & N. Torrance (Eds.), *The handbook of education and human development: New models of learning, teaching, and schooling* (pp. 225–255). Cambridge, MA: Blackwell.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, **16**, 307–322.
- Martin, J. R., Matthiessen, C., & Painter, C. (1997). *Working with functional grammar*. London: Edward Arnold.
- Matsumura, L. C., Garnier, H., Pascal, J., & Valdés, R. (2002). Measuring instructional quality in accountability systems: Classroom assignments and student achievement. *Educational Assessment*, **8**, 207–229. doi:10.1207/S15326977EA0803\_01
- Matsumura, L. C., Garnier, H., Slater, S. C., & Boston, M. B. (2008). Measuring instructional interactions “at-scale.” *Educational Assessment*, **13**, 267–300.
- Matsumura, L. C., Patthey-Chavez, G. G., Valdés, R., & Garnier, H. (2002). Teacher feedback, writing assignment quality, and third-grade students’ revision in lower- and higher-achieving urban schools. *Elementary School Journal*, **103**, 3–25.
- Monte-Sano, C. (2008). Qualities of historical writing instruction: A comparative case study of two teachers’ practices. *American Educational Research Journal*, **45**, 1045–1080.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- National Commission on Writing. (2003). *The neglected “R”: The need for a writing revolution*. Princeton, NJ: College Entrance Examination Board.
- Newmann, F. M., Bryk, A. S., & Nagaoka, J. K. (2001). *Authentic intellectual work and standardized tests: Conflict or coexistence?* Chicago: Consortium on Chicago School Research.
- Persky, H. R., Daane, M. C., & Jin, Y. (2003). The nation’s report card: Writing 2002 (IES, NCES). Retrieved from <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2003529>
- Reznitskaya, A., Anderson, R. C., McNurlen, B., Nguyen-Jahiel, K., Archodidou, A., & Kim, S. (2001). Influence of oral discussion on written argument. *Discourse Processes*, **32**, 155–175.
- Salahu-Din, D., Persky, H., & Miller, J. (2008). The nation’s report card: Writing 2007 (NCES 2008-468). Washington, DC: NCES, U.S. Department of Education.
- Sanders, T. J. M., Spooren, W. P. M., & Noordman, L. G. M. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, **15**(1), 1–35.
- Scarcella, R. (2002). Some key factors affecting English learners’ development of advanced literacy. In M. J. Schleppegrell & M. C. Colombi (Eds.), *Developing advanced literacy in first and second languages: Meaning with power* (pp. 209–226). Mahwah, NJ: Erlbaum.
- Scarcella, R. (2003). *Academic English: A conceptual framework* (University of California Linguistic Minority Research Institute Technical Report No. 2003-1). Retrieved from [http://lmri.ucsb.edu/publications/03\\_scarcella.pdf](http://lmri.ucsb.edu/publications/03_scarcella.pdf)
- Schleppegrell, M. J. (1996). Conjunction in spoken English and ESL writing. *Applied Linguistics*, **17**, 271–285.
- Schleppegrell, M. J. (2001). Linguistic features of the language of schooling. *Linguistics and Education*, **12**, 431–459.
- Schleppegrell, M. J. (2004). *The Language of schooling: A functional linguistics perspective*. Mahwah, NJ: Erlbaum.
- Schleppegrell, M. J., Achugar, M., & Oteiza, T. (2004). The grammar of history: Enhancing content-based instruction through a functional focus on language. *TESOL Quarterly*, **38**, 67–93
- Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In D. R. Olson & N. Torrance (Eds.), *Cambridge handbook of literacy*. New York: Cambridge University Press.
- Teachers of English to Speakers of Other Languages. (2009). *ESL standards for pre-K–12 students*. Retrieved from [http://www.tesol.org/s\\_tesol/secsss.asp?CID=95&DID=1565](http://www.tesol.org/s_tesol/secsss.asp?CID=95&DID=1565)